A Generalization of the Chow-Liu Algorithm and its Application to Statistical Learning http://arxiv.org/abs/1002.2240

Joe Suzuki

February 18, 2010

Learning statistical knowledge from data takes large computation. We eventually compromise between the accuracy and the time complexity of the learning algorithms by choosing its approximation to the best solution. In this paper, we address how to efficiently estimate the dependency relation among attributes values by constructing an undirected graph (a Markov network) via the Chow-Liu algorithm [1].

The original Chow-Liu algorithm approximates a probability distribution by a Dendroid distribution expressed by a tree to obtain the best solution in the sense that the Kullback-Leibler information is the smallest from the original distribution: starting with a finite set V and real numbers $\{w_{i,j}\}_{i,j\in V, i\neq j}$

1. $E := \{\}$
2. $\mathcal{E} := \{\{i,j\} | i,j \in V, i \neq j\}$
3. $\mathcal{E} := \mathcal{E} \backslash \{\{i,j\}\}$ for $\{i,j\} \in \mathcal{E}$ maximizing $w_{i,j}$
4. if $(V, E \cup \{\{i,j\}\})$ does not contain a loop, then $E := E \cup \{\{i,j\}\}$.
5. if $\mathcal{E} \neq \{\}$, then go to 3., else terminate.

As a result, a tree (V,E) with the maximum value of $\sum_{\{i,j\}\in E} w_{i,j}$ is obtained. Mutual information $I(i,j)$ of two random variables $X^{(i)}, X^{(j)}$ is used as $w_{i,j}$ in the Chow-Liu algorithm.

If the distribution is not given but samples are given, the task is estimation rather than approximation. Then, the Chow-Liu algorithm uses the maximum likelihood estimators of mutual information rather than the true mutual information values. Then, we would only choose a high fitness tree, without considering the complexity of the trees and the number of parameters: a (unconnected) forest rather than a (spanning) tree might have been closer to the true distribution. The order of selecting pairs of nodes may be different if we take into account the simplicity of the forests/trees structures.

In 1993, Suzuki [2] proposed a modified version of the Chow-Liu algorithm based on the Minimum Description Length in which the mutual information is replaced by the one minus a penalty value defined for each pair of random variables in order to consider the simplicity of the forest. The modified algorithm obtains the best forest in the sense of MDL.

However, those random variables are assumed to take finite values. This paper deals with the general case: the Chow-Liu algorithm and its modification for general random variables: both discrete and continuous attribute values may be contained in each example.

Each random variable does not need to be discrete or continuous: if the distribution function of random variable X is

$$F_X(x) = \begin{cases} 0 & x < -1 \\ \frac{1}{2} & -1 \leq x < 0 \\ \int_0^x \frac{1}{2} g(t)dt & 0 \leq x \end{cases}$$

with $\int_0^\infty g(x)dx = 1$, then there exists no f_X such that $F_X(x) = \int_{-\infty}^x f_X(t)dt$. In this paper, we define mutual information for any pair of random variables.

Furthermore, we extend the modified Chow-Liu algorithm so that the description length rather than $(-1)\times$ the likelihood is minimized for general cases. For example, if both finite and Gaussian random variables are present, then we need to evaluate the number of parameters consisting the forest.

References

[1] C. K. Chow and C. N. Liu. Approximating discrete probability distributions with dependence trees. IEEE Transactions on Information Theory, IT-14(3):462–467, May 1968.

[2] J. Suzuki "A Construction of Bayesian Networks from Databases on the MDL principle", *Uncertainty in Artificial Intelligence*, Washington DC, July 1993.

[3] J. Suzuki, "On Strong Consistency of Model Selection in Classification", *IEEE Trans. on Information Theory*, vol.52, issue 11, pp.4767-4774, Nov. 2006.